More Regression Algebra

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

More Regression Algebra

Introduction

- A number of important multivariate methods build on the algebra of multivariate linear regression, because they are least squares multiple regression systems, i.e., systems where one or more criteria are predicted as linear combinations of one or more predictors, with optimal prediction defined by a least squares criterion.

Introduction

- A number of important multivariate methods build on the algebra of multivariate linear regression, because they are least squares multiple regression systems, i.e., systems where one or more criteria are predicted as linear combinations of one or more predictors, with optimal prediction defined by a least squares criterion.
- In this module, we discuss some key results in multiple regression and multivariate regression that have significant implications in the context of other multivariate methods.

Introduction

- A number of important multivariate methods build on the algebra of multivariate linear regression, because they are least squares multiple regression systems, i.e., systems where one or more criteria are predicted as linear combinations of one or more predictors, with optimal prediction defined by a least squares criterion.
- In this module, we discuss some key results in multiple regression and multivariate regression that have significant implications in the context of other multivariate methods.
- We illustrate the algebra with a couple of theoretical derivations.

The Model

- Unlike the fixed score multiple regression model frequently employed, this one assumes that both predictor and criterion variables are random.

The Model

- Unlike the fixed score multiple regression model frequently employed, this one assumes that both predictor and criterion variables are random.
- Suppose we have a random variable y that we wish to predict from a set of random variables that are in the random vector $\mathbf{x}$.

The Model

- Unlike the fixed score multiple regression model frequently employed, this one assumes that both predictor and criterion variables are random.
- Suppose we have a random variable y that we wish to predict from a set of random variables that are in the random vector $\mathbf{x}$.
- To simplify matters, assume all variables are in deviation score form, i.e., have means of zero.

The Model

- Unlike the fixed score multiple regression model frequently employed, this one assumes that both predictor and criterion variables are random.
- Suppose we have a random variable y that we wish to predict from a set of random variables that are in the random vector $\mathbf{x}$.
- To simplify matters, assume all variables are in deviation score form, i.e., have means of zero.
- The prediction system is linear, so we may write

$$y = \boldsymbol{\beta}'\mathbf{x} + e \tag{1}$$

Solution for β Weights

- We choose β to minimize the expected squared error, i.e., to minimize $\mathrm{E}(e^2)$.

Solution for β Weights

- We choose β to minimize the expected squared error, i.e., to minimize $\mathrm{E}(e^2)$.
- It is easy to see (C.P.) that

$$\mathrm{E}(e^2) = \sigma_y^2 - 2\boldsymbol{\sigma}_{yx}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} \tag{2}$$

Solution for β Weights

- We choose β to minimize the expected squared error, i.e., to minimize $\mathrm{E}(e^2)$.
- It is easy to see (C.P.) that

$$\mathrm{E}(e^2) = \sigma_y^2 - 2\boldsymbol{\sigma}_{yx}\boldsymbol{\beta} + \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} \tag{2}$$

- Minimizing this involves taking the partial derivative of $\mathrm{E}(e^2)$ with respect to $\boldsymbol{\beta}$, setting the resulting equation to zero, and solving for $\boldsymbol{\beta}$. The well-known result is that

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xy} \tag{3}$$

Multiple Linear Regression: Solution for β Weights

- The preceding result assumed a single criterion variable y.

Multiple Linear Regression: Solution for β Weights

- The preceding result assumed a single criterion variable y.
- In least squares multivariate linear regression, we have 2 or more criteria, so the model becomes

$$\mathbf{y} = \boldsymbol{\beta}'\mathbf{x} + \mathbf{e} \tag{4}$$

Multiple Linear Regression: Solution for β Weights

- The preceding result assumed a single criterion variable y.
- In least squares multivariate linear regression, we have 2 or more criteria, so the model becomes

$$\mathbf{y} = \boldsymbol{\beta}'\mathbf{x} + \mathbf{e} \tag{4}$$

- In this case, we wish to select $\boldsymbol{\beta}$ to minimize the overall average squared error, i.e., to minimize $\operatorname{Tr} \mathrm{E}(\mathbf{ee}')$. It turns out that the solution is essentially the same as before, i.e.,

$$\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \tag{5}$$

Orthogonality Properties

Least Squares β Weights Imply Orthogonality

- Suppose we have linear regression system where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$. There are a number of immediate consequences.

Orthogonality Properties

Least Squares β Weights Imply Orthogonality

- Suppose we have linear regression system where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$. There are a number of immediate consequences.
- One consequence is that $\mathbf{x}$ and $\mathbf{e}$ are orthogonal, because their covariance matrix is a null matrix.

$$
\begin{aligned}
\text{Cov}(\mathbf{x}, \mathbf{e}) &= \text{E}(\mathbf{x}\mathbf{e}') \\
&= \text{E}(\mathbf{x}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x})') \\
&= \text{E}(\mathbf{x}\mathbf{y}') - \text{E}(\mathbf{x}\mathbf{x}'\boldsymbol{\beta}) \\
&= \boldsymbol{\Sigma}_{xy} - \boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \\
&= \boldsymbol{\Sigma}_{xy} - \mathbf{I}\boldsymbol{\Sigma}_{xy} \\
&= 0
\end{aligned}
$$

Orthogonality Properties

Least Squares β Weights Imply Orthogonality

- Suppose we have linear regression system where $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}$. There are a number of immediate consequences.
- One consequence is that **x** and **e** are orthogonal, because their covariance matrix is a null matrix.

$$
\begin{aligned}
\text{Cov}(\mathbf{x}, \mathbf{e}) &= \text{E}(\mathbf{x}\mathbf{e}') \\
&= \text{E}(\mathbf{x}(\mathbf{y} - \boldsymbol{\beta}'\mathbf{x})') \\
&= \text{E}(\mathbf{x}\mathbf{y}') - \text{E}(\mathbf{x}\mathbf{x}'\boldsymbol{\beta}) \\
&= \boldsymbol{\Sigma}_{xy} - \boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy} \\
&= \boldsymbol{\Sigma}_{xy} - \mathbf{I}\boldsymbol{\Sigma}_{xy} \\
&= 0
\end{aligned}
$$

- Of course, if **x** and **e** are orthogonal, $\hat{\mathbf{y}}$ and **e** must also be orthogonal.

Orthogonality Implies a Least Squares β

- We have seen that a least squares $\boldsymbol{\beta}$ implies orthogonality.

Orthogonality Implies a Least Squares β

- We have seen that a least squares $\boldsymbol{\beta}$ implies orthogonality.
- It turns out that, in a linear system of the form $\mathbf{y} = \boldsymbol{\beta}'\mathbf{x} + \mathbf{e}$, orthogonality of $\mathbf{x}$ and $\mathbf{e}$ implies that the $\boldsymbol{\beta}$ must be the least squares $\boldsymbol{\beta}$. (You can prove this as a homework assignment.)

Error Covariance Structure

- As a straightforward consequence of the formula for a least squares $\boldsymbol{\beta}$, the covariance matrix of the errors in least squares regression is

$$\begin{aligned}\boldsymbol{\Sigma}_{ee} &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} \\ &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\end{aligned}$$

Error Covariance Structure

- As a straightforward consequence of the formula for a least squares $\boldsymbol{\beta}$, the covariance matrix of the errors in least squares regression is

$$\begin{aligned}\boldsymbol{\Sigma}_{ee} &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta} \\ &= \boldsymbol{\Sigma}_{yy} - \boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Sigma}_{xy}\end{aligned}$$

- In this case, $\boldsymbol{\Sigma}_{ee}$ is the *partial covariance matrix* of the variables in **y**, with those in **x** partialled out.

Coefficient of Determination

- The coefficient of determination R^2_{pop} is the square of the correlation between the predicted scores and the criterion scores.

Coefficient of Determination

- The coefficient of determination R^2_{pop} is the square of the correlation between the predicted scores and the criterion scores.
- As a generalization of something we showed in Psychology 310, it is easy to prove (C.P.) that $\text{Cov}(y_j, \hat{y}_j) = \text{Var}(\hat{y}_j)$, and we shall use that fact below.

Coefficient of Determination

- The coefficient of determination R^2_{pop} is the square of the correlation between the predicted scores and the criterion scores.
- As a generalization of something we showed in Psychology 310, it is easy to prove (C.P.) that $\text{Cov}(y_j, \hat{y}_j) = \text{Var}(\hat{y}_j)$, and we shall use that fact below.
- The correlation between the jth criterion variable y_j and the predictors is given by

$$
\begin{aligned}
R_j &= \frac{\text{Cov}(y_j, \hat{y}_j)}{\sqrt{\text{Var}(y_j)\,\text{Var}(\hat{y}_j)}} \\
&= \frac{\text{Var}(\hat{y}_j)}{\sqrt{\text{Var}(y_j)\,\text{Var}(\hat{y}_j)}} \\
&= \sqrt{\frac{\text{Var}(\hat{y}_j)}{\text{Var}(y_j)}}
\end{aligned}
$$

whence

$$
R_j^2 = \frac{\text{Var}(\hat{y}_j)}{\text{Var}(y_j)}
$$

Coefficient of Determination

- We then obtain

$$
\begin{aligned}
R_j^2 &= \frac{\text{Var}(\hat{y}_j)}{\text{Var}(y_j)} \\
&= \frac{\boldsymbol{\sigma}'_{y_j\mathbf{x}} \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\sigma}_{\mathbf{x}y_j}}{\sigma^2_{y_j}} \\
&= \frac{\boldsymbol{\sigma}'_{y_j\mathbf{x}} \boldsymbol{\beta}_j}{\sigma^2_{y_j}}
\end{aligned}
$$

Additivity of Covariances

- In a least squares linear regression system, we may write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, and, because the predicted and error portions are uncorrelated, we may write
$$\text{Var}(\mathbf{y}) = \text{Var}(\hat{\mathbf{y}}) + \text{Var}(\mathbf{e}) \tag{6}$$

Additivity of Covariances

- In a least squares linear regression system, we may write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, and, because the predicted and error portions are uncorrelated, we may write
$$\text{Var}(\mathbf{y}) = \text{Var}(\hat{\mathbf{y}}) + \text{Var}(\mathbf{e}) \tag{6}$$
- Furthermore, since $\hat{\mathbf{y}} = \boldsymbol{\beta}'\mathbf{x}$, we may also write
$$\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}_{yy} = [\boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}] + [\boldsymbol{\Sigma}_{yy} - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}] \tag{7}$$

Additivity of Covariances

- In a least squares linear regression system, we may write $\mathbf{y} = \hat{\mathbf{y}} + \mathbf{e}$, and, because the predicted and error portions are uncorrelated, we may write
$$\text{Var}(\mathbf{y}) = \text{Var}(\hat{\mathbf{y}}) + \text{Var}(\mathbf{e}) \tag{6}$$
- Furthermore, since $\hat{\mathbf{y}} = \boldsymbol{\beta}'\mathbf{x}$, we may also write
$$\text{Var}(\mathbf{y}) = \boldsymbol{\Sigma}_{yy} = [\boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}] + [\boldsymbol{\Sigma}_{yy} - \boldsymbol{\beta}'\boldsymbol{\Sigma}_{xx}\boldsymbol{\beta}] \tag{7}$$
- This formula gives explicit expressions for the partitioning of variances and covariances into predicted and error components in least squares multivariate regression.

Applications

- In this section, we examine a few well-known applications of the theory developed in previous sections.

Applications

Regression Component Analysis

- "Component analysis" is a well-known alternative to common factor analysis.

Applications

Regression Component Analysis

- "Component analysis" is a well-known alternative to common factor analysis.
- Both component and factor analysis are commonly thought of as "factor analytic methods," although they have some important differences.

Applications

Regression Component Analysis

- "Component analysis" is a well-known alternative to common factor analysis.
- Both component and factor analysis are commonly thought of as "factor analytic methods," although they have some important differences.
- The best known example of component analysis is Principal Component Analysis, or PCA.

Applications

Regression Component Analysis

- "Component analysis" is a well-known alternative to common factor analysis.
- Both component and factor analysis are commonly thought of as "factor analytic methods," although they have some important differences.
- The best known example of component analysis is Principal Component Analysis, or PCA.
- PCA is a special case of a more general system known as "regression component analysis" (Schönemann and Steiger, 1976).

Applications

Regression Component Analysis

- A set of "components" $\mathbf{x}$ of a set of random variables $\mathbf{y}$ is any set of linear combinations of $\mathbf{y}$.

Applications

Regression Component Analysis

- A set of "components" $\mathbf{x}$ of a set of random variables $\mathbf{y}$ is any set of linear combinations of $\mathbf{y}$.
- Specifically, we write

$$\mathbf{x} = \mathbf{B}'\mathbf{y} \tag{8}$$

Applications

Regression Component Analysis

- A set of "components" $\mathbf{x}$ of a set of random variables $\mathbf{y}$ is any set of linear combinations of $\mathbf{y}$.
- Specifically, we write

$$\mathbf{x} = \mathbf{B}'\mathbf{y} \tag{8}$$

- A regression component system is of the form

$$\mathbf{y} = \mathbf{F}\mathbf{x} + \mathbf{e} \tag{9}$$

where $\mathbf{x} = \mathbf{B}'\mathbf{y}$ is a set of components of $\mathbf{y}$, and $\mathbf{F}$, known as the "component pattern", is the set of least squares linear regression weights for predicting $\mathbf{y}$ from $\mathbf{x}$.

Applications

Regression Component Analysis

- Notice that the system is completely tautological in one sense, since $\mathbf{e} = (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y}$, and so of course

$$\mathbf{y} = \mathbf{F}(\mathbf{B}'\mathbf{y}) + (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y} \tag{10}$$

Applications

Regression Component Analysis

- Notice that the system is completely tautological in one sense, since $\mathbf{e} = (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y}$, and so of course

$$\mathbf{y} = \mathbf{F}(\mathbf{B}'\mathbf{y}) + (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y} \tag{10}$$

- Once $\mathbf{B}$ is established for a given $\mathbf{y}$, the components are uniquely defined. In a sense, examining $\mathbf{B}$ establishes the relationship between the components and the variables used to construct them.

Applications

Regression Component Analysis

- Notice that the system is completely tautological in one sense, since $\mathbf{e} = (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y}$, and so of course

$$\mathbf{y} = \mathbf{F}(\mathbf{B}'\mathbf{y}) + (\mathbf{I} - \mathbf{F}\mathbf{B}')\mathbf{y} \tag{10}$$

- Once $\mathbf{B}$ is established for a given $\mathbf{y}$, the components are uniquely defined. In a sense, examining $\mathbf{B}$ establishes the relationship between the components and the variables used to construct them.
- The real "payoff" for RCA is when the $p \times m$ matrix $\mathbf{B}'$ has only a few columns, so that p, the number of variables in $\mathbf{y}$, is much smaller than m, the number of components, and yet the error variance is small.

Applications

Regression Component Analysis

- In a regression component system, once **B** is defined, then for any set of data, the "component pattern" **F** is automatically defined. Conversely, any given **F** corresponds to a derivable **B**.

Applications

Regression Component Analysis

- In a regression component system, once $\mathbf{B}$ is defined, then for any set of data, the "component pattern" $\mathbf{F}$ is automatically defined. Conversely, any given $\mathbf{F}$ corresponds to a derivable $\mathbf{B}$.
- As an example, suppose we try to derive the facts about $\mathbf{F}$ and $\mathbf{B}$.

Applications

Regression Component Analysis

- In a regression component system, once $\mathbf{B}$ is defined, then for any set of data, the "component pattern" $\mathbf{F}$ is automatically defined. Conversely, any given $\mathbf{F}$ corresponds to a derivable $\mathbf{B}$.
- As an example, suppose we try to derive the facts about $\mathbf{F}$ and $\mathbf{B}$.
- To begin with, suppose that the scores in $\mathbf{y}$ are in deviation score form. Since $\mathbf{x}$, $\mathbf{e}$, and $\hat{\mathbf{y}}$ are all linear combinations of $\mathbf{y}$, they must also in deviation score form.

Applications

Regression Component Analysis

- In a regression component system, once $\mathbf{B}$ is defined, then for any set of data, the "component pattern" $\mathbf{F}$ is automatically defined. Conversely, any given $\mathbf{F}$ corresponds to a derivable $\mathbf{B}$.
- As an example, suppose we try to derive the facts about $\mathbf{F}$ and $\mathbf{B}$.
- To begin with, suppose that the scores in $\mathbf{y}$ are in deviation score form. Since $\mathbf{x}$, $\mathbf{e}$, and $\hat{\mathbf{y}}$ are all linear combinations of $\mathbf{y}$, they must also in deviation score form.
- To begin with, let me ask you to derive $\mathbf{\Sigma}_{yx}$, the covariance matrix between $\mathbf{y}$ and the components in $\mathbf{x}$, in terms of $\mathbf{\Sigma}_{yy}$ and $\mathbf{B}$.

Applications

Regression Component Analysis

- In a regression component system, once $\mathbf{B}$ is defined, then for any set of data, the "component pattern" $\mathbf{F}$ is automatically defined. Conversely, any given $\mathbf{F}$ corresponds to a derivable $\mathbf{B}$.
- As an example, suppose we try to derive the facts about $\mathbf{F}$ and $\mathbf{B}$.
- To begin with, suppose that the scores in $\mathbf{y}$ are in deviation score form. Since $\mathbf{x}$, $\mathbf{e}$, and $\hat{\mathbf{y}}$ are all linear combinations of $\mathbf{y}$, they must also in deviation score form.
- To begin with, let me ask you to derive $\boldsymbol{\Sigma}_{yx}$, the covariance matrix between $\mathbf{y}$ and the components in $\mathbf{x}$, in terms of $\boldsymbol{\Sigma}_{yy}$ and $\mathbf{B}$.
- Before clicking on the button to move to the next slide, take a few seconds to see if you can derive the answer. (Hint: $\boldsymbol{\Sigma}_{yx} = \mathrm{E}(\mathbf{y}\mathbf{x}')$.)

Applications

Regression Component Analysis

- Here is the solution.

$$\boldsymbol{\Sigma}_{yx} = \mathrm{E}(\mathbf{y}\mathbf{x}') \tag{11}$$

Applications

Regression Component Analysis

- Here is the solution.

$$\boldsymbol{\Sigma}_{yx} = \mathrm{E}(\mathbf{y}\mathbf{x}') \tag{11}$$

- But $\mathbf{x} = \mathbf{B}'\mathbf{y}$, so

$$\begin{aligned}
\boldsymbol{\Sigma}_{yx} &= \mathrm{E}(\mathbf{y}(\mathbf{B}'\mathbf{y})') \\
&= \mathrm{E}(\mathbf{y}\mathbf{y}'\mathbf{B}) \\
&= \mathrm{E}(\mathbf{y}\mathbf{y}')\mathbf{B} \\
&= \boldsymbol{\Sigma}_{yy}\mathbf{B}
\end{aligned} \tag{12}$$

Applications

Regression Component Analysis

- Here is another fairly straightforward problem for you.

Applications

Regression Component Analysis

- Here is another fairly straightforward problem for you.
- Express $\boldsymbol{\Sigma}_{xx}$, the variance-covariance matrix of the $\mathbf{x}$ components, in terms of $\mathbf{B}$ and $\boldsymbol{\Sigma}_{yy}$, the variance-covariance matrix of the variables in $\mathbf{y}$. When you have your answer, click on the button to move on to the next slide.

Applications

Regression Component Analysis

- Here is the solution.

Applications

Regression Component Analysis

- Here is the solution.
-

$$
\begin{aligned}
\boldsymbol{\Sigma}_{xx} &= \mathrm{E}(\mathbf{x}\mathbf{x}') \\
&= \mathrm{E}(\mathbf{B}'\mathbf{y}\mathbf{y}'\mathbf{B}) \\
&= \mathbf{B}'\,\mathrm{E}(\mathbf{y}\mathbf{y}')\mathbf{B} \\
&= \mathbf{B}'\boldsymbol{\Sigma}_{yy}\mathbf{B} \qquad (13)
\end{aligned}
$$

Applications

Regression Component Analysis

- Finally, show how to construct a formula for computing $\mathbf{F}$, the component pattern, from $\mathbf{B}$ and $\boldsymbol{\Sigma}_{yy}$.

Applications

Regression Component Analysis

- Finally, show how to construct a formula for computing $\mathbf{F}$, the component pattern, from $\mathbf{B}$ and $\boldsymbol{\Sigma}_{yy}$.
- Hint: remember that in a regression system, the linear weights β' for predicting $\mathbf{y}$ from $\mathbf{x}$ are computed as $\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$.

Applications

Regression Component Analysis

- Finally, show how to construct a formula for computing $\mathbf{F}$, the component pattern, from $\mathbf{B}$ and $\boldsymbol{\Sigma}_{yy}$.
- Hint: remember that in a regression system, the linear weights β' for predicting $\mathbf{y}$ from $\mathbf{x}$ are computed as $\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_{xx}^{-1}$.
- When you have your answer, click on the button to continue on to the next slide.

Applications

Regression Component Analysis

- The solution is as follows. In this context, we have already established that $\mathbf{\Sigma}_{yx} = \mathbf{\Sigma}_{yy}\mathbf{B}$, and that $\mathbf{\Sigma}_{xx} = \mathbf{B}'\mathbf{\Sigma}_{yy}\mathbf{B}$.

Applications

Regression Component Analysis

- The solution is as follows. In this context, we have already established that $\mathbf{\Sigma}_{yx} = \mathbf{\Sigma}_{yy}\mathbf{B}$, and that $\mathbf{\Sigma}_{xx} = \mathbf{B}'\mathbf{\Sigma}_{yy}\mathbf{B}$.
- In a regression component system, $\mathbf{F}$ plays the same role as β' in the general multivariate linear regression model. So

$$\begin{aligned} \mathbf{F} &= \mathbf{\Sigma}_{yx}\mathbf{\Sigma}_{xx}^{-1} \\ &= \mathbf{\Sigma}_{yy}\mathbf{B}(\mathbf{B}'\mathbf{\Sigma}_{yy}\mathbf{B})^{-1} \end{aligned} \tag{14}$$